

Dr. Marshall Nirenberg
Genetic Versus Neural Information Processing Systems

I'd like to talk very, very informally and do something that a biochemist should never do, does always in private and never publically if talking and speculate about something which one doesn't know very much about. I do this all the time with friends, especially in discussions but not in public so I want to emphasize that, rather than misleading anybody. I spend most of my time thinking about genetic information processing, mechanisms, and so forth and only recently have I really begun thinking seriously about information processing by neurons and so I just wish to say that it may sound a little rash to speculate, but, at any rate, if one views a simple cell, such as E. coli, I guess E. coli, one sees the flow of materials and interconversion of molecules of various sorts, a flow of energy and also a flow of information. Now the information processing machinery in rapidly growing bacteria will account for 50 or 60% of the total dry weight of the cell. The information is encoded, as you know, in a special class of molecules. It's memory has a specific address and DNA is a very long macro-molecule consisting of only 4 kinds of characters in repeating sequence. A linear string of characters. In E. coli the information may be approximately 3 million letters in length in bacterial chromosomes. A human cell with DNA would consist of approximately 1,000 or 1,500 times as much information. The sequence of the letters in DNA correspond to the sequence of amino-acids in protein and as you know there are 20 amino-acids in protein, 4 bases in DNA and the information in DNA is decoded as a linear form, a one dimensional form and it's decoded by the reading apparatus starting at one point

in reading a word sequentially from left to right. The average protein contains some 400 amino acids in length so bacteria may contain about 3,000 kinds of proteins, probably maximally, and for each chromosome there is approximately a million and a half molecules with protein in bacteria so the number of molecules in each kind of protein may vary quite considerably from one molecule up to 100,000 or more molecules. The properties of the template, of the DNA template or an RNA template are really very, very simple. The nucleic-acid sequence, base sequence specifies a kind of object; one molecule of a particular kind. It differentiates, it selects an example of that kind of molecule from many other types of molecules. It specifies the relative position of the molecule, that is relative to the previous molecule it selected and it also specifies the time of the event, relative to the previous event that occurred. So the template really serves both as a template for other molecules and as a biological clock because the words are read in sequential fashion. The principle really is a Turing machine principle. That is in the 30's you say British mathematician, Charles Turing devised a very simple kind of machine that would have a reading head that was attached to a tape. The program would travel along the tape and would follow only one instruction at a time and it could read only one space at a time. Basically, this is the same principle employed in all of genetic coding processing and it has some really unique characteristics, for example, one can program one kind of reading machine, Turing machine, to perform any kind of task that any other Turing machine can be programmed for. One of the major principles is that there are relatively few kinds of units, and that great diversity

is achieved, and also great complexity is obtained simply by varying the sequence of the units. The units are rather standard ones. I should emphasize that all the information available today indicates that the language is largely a universal language, that all species employ essentially almost the same genetic language. I think that this is important when one considers the origin and the evolution of the language and also, considering the problem of neural memory, ^{i.e.} information processing by the nervous system. The origin of the code is an interesting question. One may ask why is the code set up the way it is? Is there any rational reason for it, or is it the result of a unique set of circumstances that happened only once in the dim and distant past? Then after this unique set of circumstances the code that we're aware of may have evolved from this but the original event may still have control of the format or the form of the code as we see it today. The oldest fossils that are known are micro-organisms. They've been estimated to be 3 billion years of age, found by Barcorn and his colleagues. Fossil micro-organisms 2 billion and 1 billion years old have also been reported but the first abundant fossils are found only about 600 million years ago. About 500 million years ago all the invertebrate phyla had evolved and the first vertebrates were formed and mammals rather soon after the code evolved, that the code was probably fixed, the language was fixed because once a sufficient amount of information had been put into the system to form a bacterium, I think that probably it would be difficult to make terribly major changes in the code without destroying the information that

had already been put into the system unless it was possible, simultaneously to alter the translation apparatus so that the information that's stored can be retrieved. I think this is a fairly reasonable argument and one may well apply this same type of thinking; I wonder if it can be applied at least, to the problem of information processing by the nervous system. There are several differences, I think obvious differences: First, the genetic code must have evolved before any neural code could have evolved. The genetic code probably evolved as the first primitive cells evolved, perhaps between 600 million and 3 billion years ago. When the neural code or neural information processing mechanisms must have evolved at a later date when cells were more adept. They were probably quite sophisticated chemically during the transition, perhaps between single cells and multi-cellular forms of life. And this has another consequence, that is, that almost surely, I think, the basic mechanisms for processing neural information undoubtedly were selected from a large population of precursors or precursor mechanisms whereas one doesn't really know. If this is true in a case of genetic code it could have been one extremely rare set of circumstances. Probably the basic mechanisms for processing neural information had evolved and perhaps were probably fixed by 500 million years ago. All of the information that is available that I can see, at least that I'm aware of, which is not very much, but in regard certainly to the neural information processing mechanisms, but from all the available evidence that I see indicates rather strongly that universal mechanisms are at work, that the neural transmitter substances, neural hormones one finds in very simple forms and pretty much the same substances are found in flatworms as one finds in mammals. And similar kinds of mechanisms; synaptic vesicles and nerve endings and

so forth. So I would think that the basic mechanisms, there are relatively few basic mechanisms, operative, that they must be systematic mechanisms, and quite simple and logical mechanisms, and that probably they're almost universal similar to the genetic code and that after enough information had been put into this system, into the nervous system, in terms of basic mechanisms, that it's difficult; one can add on new mechanisms, yes, but to change old mechanisms would be difficult because you destroy all stored information if you did do this. I suppose one of the most important aspects of living organisms is the reliability of the mechanisms, of the machinery, and the accuracy also of the machinery. It's really remarkable if one considers that the average protein is 400 amino-acids long, in any series of sequential calculations obviously the accuracy decreases quite markedly as the number of serial steps increases. The accuracy of protein synthesis, from everything that we know is usually quite high and I think that the basic reason for this is, well, there are two basic strategies, I think, that the cell employs to enhance the accuracy of protein synthesis: First, redundancy is a logical kind of redundancy, alternate words, in most cases, have identical first and second letters but only alternate, vary, the third letters so it's systematic redundancy. Secondly, one must consider the number of serial steps and the number of parallel steps here. Protein synthesis is largely a parallel operation. There are serial steps, but interestingly the serial steps enhance the accuracy of protein synthesis. There's one very, very interesting example of this that was reported by Berg and Norris. It was an error-correcting mechanism. As you

are probably well aware, the amino-acids are activated and the cell contains enzymes/^{that}are specific to each amino-acid, different enzymes, one for each amino acid. The enzyme is a device that selects a particular, appropriate species of amino acid and the appropriate species of transfer RNA, the adapter molecule and also ATP. It catalyzes first the formation of an intermediate, an activated amino acid. Then the same enzyme will select a species of transfer RNA and it will transfer the amino acid from the activated state that's in to the transfer RNA. Berg and Norris found that a mistake could be made in the first step, the enzyme would make a mistake and it would recognize the wrong amino acid. In the second step it would correct this error by hydrolyzing, by breaking down this erroneous intermediate. It would not catalyze the transfer of the wrong amino acid to the correct tRNA. This is a beautiful error-correcting mechanism and a perfect example of a mechanism whereby two serial steps enhance the accuracy rather than decrease the accuracy of protein synthesis. Although the other steps are sequential steps, in almost all cases the accuracy of selecting one amino acid or adapter molecule with an amino acid linked to it is independent of the previous selections so an error in one selection usually does not affect the accuracy of further selections. I think that this is the crux of the strategy, right here. There are certainly places where the accuracy of selecting one amino acid does affect the accuracy of selecting subsequent amino acids. This is in initiation, in starting. If the first three bases are incorrectly selected, incorrectly phased, then all subsequent reading will obviously be out of phase and will be erroneous. Secondly, if an error is made and a word that corresponds to stop,

terminate or a protein synthesis is terminated because a word is recognized incorrectly, obviously subsequent selections, there will be no reading, so this will affect it. But, in all other cases that I'm aware of an error of selecting one amino acid will not affect the accuracy of subsequent amino acids. The efficiency of this system is quite remarkable when one considers that the average rate of reading of the reading head, the ribosome along the tape, is relatively slow. The rate of reading is somewhat faster than the average person reads, perhaps in the neighborhood of 1,000 words a minute are read. The efficiency is enhanced very greatly because the cell contains many ribosomes, many reading heads, about 15,000 sites for protein synthesis per bacterial chromosome and one message can be read simultaneously at many different sites, by many different ribosomes. In a relatively short time a great many words are translated. E. coli have a generation time, under optimum conditions, of 20 to 25 minutes. During this time, 25 minutes, about 500 million amino acids are incorporated to protein. So it's a relatively efficient process. We come to the question, really, of learning and of memory. As I said, in the genetic memory, memory has a specific address, corresponding to each memory is a specific molecule, I don't know what the answer is, obviously, when applied to the nervous system. In recent years there has been a great deal of speculation and extrapolation that it may be each neural memory corresponds to a specific molecule of RNA or a kind of molecule of RNA. I don't know, as I say, the answer, but my bias is probably against this possibility although I think it's too early, really to evaluate. I doubt if there's sufficient information available to judge. All cells require nucleic

acid, have protein synthesis. It's possible, for example, that a specific permease of some sort can be induced, that one kind or just a limited number of kinds of molecules, proteins must be synthesized to facilitate memory, to lay down the memory traits. It's clear that a chemical reaction of some sort, I would think it highly likely, must take place. The nature of the chemical reaction remains to be defined.

In genetics, for many years there was a great deal of confusion about the various theories of learning. One can think of learning on the genetic level, memory, of immunologic memory, and memory in the nervous system. The two basic theories were the instructive theory and the selective theory. By instructive theory one means that something must be present. I think that the easiest thing to do is just to give an example of it. This was clarified by a very famous experiment by Luria and Delbruck during the '40's. This experiment was to take a petri dish, to put some E. coli on this and to grow up a uniform lawn of the E. coli on the petri dish. It was known that if in a flask, for example, in a liquid medium that if one adds an antibiotic streptomycin to a flask, that it will initially kill most of the bacteria, the next morning one sees resistant population of bacteria. So the question really was, is the streptomycin necessary for the apparent mutation that took place? In a very clever and simple experiment they showed that the mutation would not require, that rather the mutation was independent of the streptomycin. Streptomycin was not required for this. They grew up a lawn of streptomycin sensitive to E. coli. Then on two plates here which contain streptomycin - no streptomycin in this plate, then by pressing a velvet cloth on this and then going directly here and directly here one would transfer colonies from

here to here to here, and here. And the orientation of colonies would be similar. One found on a plate like this that one little colony, a clone, the descendents of a single cell would grow, let's say here, here, and here. Then on this plate, with the same orientation, one found exactly the same orientation of the three bacteria that had spontaneously become resistant to streptomycin in the absence of streptomycin. It was very clear that the resistant bacteria were already there and were simply selected for by the streptomycin. This is a very simple experiment but it's almost precisely the same kind of situation that can be found in the antibody production mechanism. There seems to be cells that are present. Their descendents are selected for and the antibody production is due to the selection of the population of cells. One wonders, really, whether learning is due to selection process also as it is in genetic information processing and immunologic information processing. The philosopher, Locke, thought of the mind as a completely blank slate that experience writes on but Socrates thought that you can't learn anything that's not already in the mind. This is very similar to a selective type of hypothesis, that one simply selects out from something that's already there. The reason he thought this was because by questions only one can teach. I think that these are problems that one should think about. Thank you very much.